Week 12 Homework

1. **Non-convex Gradient Descent** (8 points)

Consider minimizing a differentiable function f with $\mathrm{dom}(f) = \mathbb{R}^n$, whose gradient is L-Lipschitz continuous for a constant $L > 0$, meaning

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2, \text{ for all } x, y.$$

We will run gradient descent, starting from x_0, with the updates

$$x_{k+1} = x_k - t \cdot \nabla f(x_k),$$

where $t \leq 1/L$. (Notice that we assume nothing about convexity of f.)

(1) Prove that

$$f(x_{k+1}) \leq f(x_k) - \left(1 - \frac{Lt}{2}\right) t\|\nabla f(x_k)\|_2^2.$$

(2) Use $t \leq 1/L$, and rearrange the previous result, to get

$$\|\nabla f(x_k)\|_2^2 \leq \frac{2}{t}\left(f(x_k) - f(x_{k+1})\right).$$

(3) Sum the previous result over all iterations from $0, \ldots, k$ to establish

$$\sum_{i=0}^{k} \|\nabla f(x_i)\|_2^2 \leq \frac{2}{t}\left(f(x_0) - f(x^*)\right).$$

(4) Lower bound the sum in the previous result to get

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\|_2 \leq \sqrt{\frac{2}{t(k+1)}\left(f(x_0) - f(x^*)\right)}.$$

As a result, we have proved that gradient descent reaches an ϵ-substationary point x, such that $\|\nabla f(x)\|_2 \leq \epsilon$, in $O(1/\epsilon^2)$ iterations.

Hint: You may use here that (need to clarify when using)

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|_2^2, \quad \text{for all } x, y.$$

2. **Ill-conditioned linear equality** (8 points)

Consider the problem of solving $Ax = b$, where A is a ill-conditioned matrix. This can be converted into an optimization problem:

$$\min_x \frac{1}{2}\|Ax - b\|_2^2$$

Define $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, we could instead consider the corresponding proximal operator

$$\mathrm{Prox}_{\alpha f}(v) = \arg\min_x \alpha f(x) + \frac{1}{2}\|x - v\|^2.$$

(1) Derive the exact form of iteration $x_{k+1} = \mathrm{Prox}_{\alpha f}(x_k)$.

(2) Achieve the iteration by code with matrix A, b given in file `1A.csv` and `1b.csv`. Choose your own α to make this iteration numerical stable and efficient. Stop when $f(x_k) \le 2 \cdot 10^{-2}$. Plot the corresponding figure of $\log(k)$ vs $\log(\|x_k - x^*\|_2)$ and $\log(k)$ vs $\log(f(x_k))$.

3. **LASSO** (9 points)

Consider the following LASSO problem:

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \|x\|_1.$$

The matrix A, b is provided in file `2A.csv, 2b.csv`.

(1) Achieve the sub-gradient method with code:

$$x_{k+1} = x_k - \alpha_k g_k,\ g_k \in \partial h(x),$$

where $h(x) = \frac{1}{2}\|Ax - b\|_2^2 + \|x\|_1$. Choose $\alpha_k = c \cdot k^{-\beta}$ and $c = 0.01, \beta = 0.5$. Start with x_0 given in file `2x0.csv` (stored as row vector $x_0^\top$). Stop when $\|x_{k+1} - x_k\| < 10^{-8}$. Plot the corresponding figure of $\log(k)$ vs $\log(\|x_{k+1} - x_k\|_2)$, $\log(k)$ vs $\log(\|x_k - x^*\|_2)$ and $\log(k)$ vs $\log(f(x_k))$.

(2) When A is full column-rank, the function $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ is strongly convex with respect to x and for all x, y we have

$$f(y) \ge f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2}\|y - x\|_2. \tag{1}$$

Given A, b as in files `2A.csv, 2b.csv`, find the maximal m that (1) is satisfied. Let $\alpha_k = \frac{1}{mk}$ and achieve the sub-gradient method. Start with x_0 given in file `2x0.csv` (stored as row vector $x_0^\top$). Stop when $\|x_{k+1} - x_k\| < 10^{-8}$. Plot the corresponding figure of $\log(k)$ vs $\log(\|x_{k+1} - x_k\|_2)$, $\log(k)$ vs $\log(\|x_k - x^*\|_2)$ and $\log(k)$ vs $\log(f(x_k))$.

1 作业说明

- 第 1 题需要理论证明，第 2,3 题需要编程报告，包含计算结果/图像/及其分析。报告提交电子版，和代码一起打包提交至网络学堂。提交作业时文件夹中应包含数据文件，保证程序可以直接在文件夹中运行。
- 编程语言不限，第 2,3 题请使用文件夹中提供的数据。
- 计分方式：第一题每小问 2 分。第二题第 (1) 问 2 分，第 (2) 问共 6 分，程序过程 2 分，两张图每张 2 分。第三题第 (1) 问 4 分，程序过程 1 分，三张图每张 1 分。第 (2) 问 5 分，算 m 过程 2 分，三张图每张 1 分。
- 请大家在截止日期前提交作业，过期不候。